Data association by loopy belief propagation

Jason L. Williams[1] **and Roslyn A. Lau**[1,2]
[1]Intelligence, Surveillance and Reconnaissance Division, DSTO, Australia
[2]Statistical Machine Learning Group, NICTA, Australia
jason.williams@dsto.defence.gov.au, roslyn.lau@dsto.defence.gov.au

Abstract − *Data association, or determining correspondence between targets and measurements, is a very difficult problem that is of great practical importance. In this paper we formulate the classical multi-target data association problem as a graphical model and demonstrate the remarkable performance that approximate inference methods, specifically loopy belief propagation, can provide. We apply it to calculating marginal association weights (e.g., for JPDA) for single scan and multiple scan problems, and to calculating a MAP hypothesis (i.e., multi-dimensional assignment). Through computational experiments involving challenging problems, we demonstrate the remarkable performance of this very simple, polynomial time algorithm; e.g., errors of less than 0.026 in marginal association weights and finding the optimal 5D assignment 99.4% of the time for a problem with realistic parameters. Impressively, the formulation commits smaller errors in association weights in challenging environments, i.e., in problems with low P_d and/or high false alarm rates. Our formulation paves the way for the expanding literature on approximate inference methods in graphical models to be applied to classical data association problems.*

Keywords: Data association, JPDA, graphical models, loopy belief propagation, multi-dimensional assignment

1 Introduction

In recent years, graphical models have emerged as a powerful tool for inference and learning in large scale systems. The promise of graphical models in tracking problems was demonstrated in [1, 2, 3]. The formulation in the former focussed on sensor networks, in which each sensor had a narrow field of view. Non-overlapping regions were defined and association variables were instantiated to hypothesise joint association events for all targets and measurements within a region. In the present study, we consider the classical data association problem, in which a single sensor surveils a large number of targets. Each target may give rise to at most one measurement, and each measurement is related to at most one target. In a sense, the resulting method decomposes the formulation of [2], such that each measurement is within its own region.

We focus on approximate solutions to two core problems in data association: firstly, calculating marginal association probabilities such as those used in Joint Probabilistic Data Association (JPDA) [4], and secondly, finding a maximum a posteriori (MAP) association configuration, similar to that sought in [5]. After introducing our notation and the tools of graphical models in Section 2, we provide (in Section 3) two graphical model formulations of the single scan, single sensor data association problem, and demonstrate the remarkable performance that is achieved in Section 4.1. Our method may be easily extended to problems involving multiple sensors and multiple time steps; we sketch this development in Section 3.3, and demonstrate results in Section 4.2.

2 Background

2.1 Data association model

We now describe this classical model, introducing our notation as we proceed. We denote by n_t the number of targets under track. While this is assumed fixed (and known), varying (and uncertain) numbers of targets can be accommodated easily through a framework such as [6]. We denote by $\boldsymbol{x}_t^i \in \mathbb{R}^n$, $i \in \{1, \ldots, n_t\}$, the state of the i-th target at time t, and by $\boldsymbol{X}_t = (\boldsymbol{x}_t^1, \ldots, \boldsymbol{x}_t^{n_t})$ the augmented state of all targets at time t.

Our measurement model hypothesises a set of measurements comprised of possible target detections and false alarms. We denote by m_t the number of measurements at time t, and by $\boldsymbol{z}_t^j \in \mathbb{R}^m$, $j \in \{1, \ldots, m_t\}$, the value of the j-th measurement. A target in state $\boldsymbol{x}_t^i$ will be detected with probability of detection $P_{\mathrm{d}}(\boldsymbol{x}_t^i)$. The measurement resulting from this detection is distributed according to the PDF $p(\boldsymbol{z}|\boldsymbol{x}_t^i)$ and is independent of all other measurements and targets.

False alarms occur according to a non-homogeneous Poisson point process with intensity $\lambda_{\mathrm{fa}}(\boldsymbol{z})$. False alarm measurements are independent of targets and their measurements. The complete set of measurements at time t is denoted as $\boldsymbol{Z}_t = (m_t, \boldsymbol{z}_t^1, \ldots, \boldsymbol{z}_t^{m_t})$. The ordering of measurements within this vector is arbitrary.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUL 2010		**00-00-2010 to 00-00-2010**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Data association by loopy belief propagation	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Defence Science and Technology Organisation (DSTO),Intelligence, Surveillance and Reconnaissance Division,Australia,**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

Presented at the 13th International Conference on Information Fusion held in Edinburgh, UK on 26-29 July 2010. Sponsored in part by Office of Naval Research, Office of Naval Research Global, and U.S. Army Research Laboratory's Army Research Office (ARO). U.S. Government or Federal Rights License.

14. ABSTRACT

Data association, or determining correspondence between targets and measurements, is a very difficult problem that is of great practical importance. In this paper we formulate the classical multi-target data association problem as a graphical model and demonstrate the remarkable performance that approximate inference methods, specifically loopy belief propagation can provide. We apply it to calculating marginal association weights (e.g., for JPDA) for single scan and multiple scan problems, and to calculating a MAP hypothesis (i.e., multi-dimensional assignment). Through computational experiments involving challenging problems we demonstrate the remarkable performance of this very simple, polynomial time algorithm; e.g., errors of less than 0.026 in marginal association weights and finding the optimal 5D assignment 99.4% of the time for a problem with realistic parameters. Impressively the formulation commits smaller errors in association weights in challenging environments, i.e., in problems with low Pd and/or high false alarm rates. Our formulation paves the way for the expanding literature on approximate inference methods in graphical models to be applied to classical data association problems.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**8**	

The complete set of measurements up to and including time t is denoted as $\boldsymbol{Z}^t = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_t)$.

The relationship between targets and measurements is described via a set of association variables. The association variables comprise of either or both of the following:

1. For each target $i \in \{1, \ldots, n_t\}$, an association variable $a_t^i \in \{0, 1, \ldots, m_t\}$, the value of which is an index to the measurement with which the target is hypothesised to be associated (zero if the target is hypothesised to have not been detected)

2. For each measurement $j \in \{1, \ldots, m_t\}$, an association variable $b_t^j \in \{0, 1, \ldots, n_t\}$, the value of which is an index to the target with which the measurement is hypothesised to be associated (zero if the measurement is hypothesised to be a false alarm)

Note that the two sets of association variables are entirely redundant: given the information from either set, the other can be reconstructed perfectly. We consider two formulations; the first, involving only target association variables $(a_t^1, \ldots, a_t^{n_t})$, we refer to as a *target oriented* formulation; the second, which incorporates both sets of variables, we refer to as a *hybrid* formulation. A similar *measurement oriented* formulation has been considered but is omitted here due to space limitations; it is closely related to the well-known Probabilistic Multiple Hypothesis Tracker (PMHT) [7].

Through straight-forward and well-known manipulations, the joint probability of measurements and associations conditioned on target states can be written as:[1]

$$p(\boldsymbol{Z}_t, \boldsymbol{a}_t | \boldsymbol{X}_t) \propto \left[\prod_{i=1}^{n_t} \left(\frac{P_d(\boldsymbol{x}_t^i) p(\boldsymbol{z}_t^{a_t^i} | \boldsymbol{x}_t^i)}{\lambda_{\text{fa}}(\boldsymbol{z}_t^{a_t^i})} \right)^{\theta_d(a_t^i)} \cdot \cdot (1 - P_d(\boldsymbol{x}_t^i))^{1-\theta_d(a_t^i)} \right] \psi_c(\boldsymbol{a}_t) \quad (1)$$

where $\boldsymbol{a}_t = (a_t^1, \ldots, a_t^{n_t})$, $\theta_d(a_t^i)$ is the detection flag:

$$\theta_d(a_t^i) = \begin{cases} 0, & a_t^i = 0 \\ 1, & a_t^i \neq 0 \end{cases}$$

and $\psi_c(\boldsymbol{a}_t)$ is the global consistency constraint

$$\psi_c(\boldsymbol{a}_t) = \begin{cases} 0, & \exists\, i, j \in \{1, \ldots, n_t\} \text{ s.t. } a_t^i = a_t^j \neq 0 \\ 1, & \text{otherwise} \end{cases}$$

The proportionality in Eq. (1) is with respect to all parameters other than the number of observations and observation values (both of which are constants at runtime).

[1] Abusing notation, we set $f(z^a)^0 = 1$ even when z^a is undefined (*i.e.*, when $a = 0$).

2.2 Graphical models

Graphical models [8, 9, 10] aim to represent and manipulate the joint probability distributions of many variables efficiently by exploiting factorisation. The Kalman filter [11] and the hidden Markov model (HMM) [12] are two examples of algorithms that exploit sparsity of a particular kind (*i.e.*, a Markov chain) to efficiently conduct inference on systems involving many probability variables. Inference methods based on the graphical model framework generalise these algorithms to a wider variety of state spaces and dependency structures.

Graphical model methods have been developed for undirected graphical models (Markov random fields), directed graphical models (Bayes nets) and factor graphs. We prefer factor graphs [13] as they provide the most expressive form of characterisation of the dependency structure. In the general case, we have nodes (*i.e.*, random variables) $n \in \mathcal{N}$, and factors (*i.e.*, dependencies) $f \in \mathcal{F}$, where each f is equipped with a set of neighbouring nodes, $\eta_f \subseteq \mathcal{N}$. The joint distribution is factored as:

$$p(x_{\mathcal{N}}) \propto \prod_{n \in \mathcal{N}} \psi_n(x_n) \prod_{f \in \mathcal{F}} \psi_f(x_{\eta_f})$$

The functions $\psi_{\cdot}(\cdot)$ collectively represent the joint probability distribution. For example, a Markov chain involving variables $(x_1, \ldots, x_n)$ may be formulated using a factor $\psi_1(x_1) = p(x_1)$ for the initial prior, and factors $\psi_{k-1,k}(x_{k-1}, x_k) = p(x_k | x_{k-1})$, $k \in \{2, \ldots, n\}$ representing the Markov transition kernels, although many other formulations are possible. Optimal inference can be conducted on tree-structured factor graphs using belief propagation (BP). BP proceeds by passing messages between nodes and factors. We denote by $\mu_{n \to f}(x_n)$ the message sent from node $n \in \eta_f$ to factor f, by $\mu_{f \to n}(x_n)$ the message sent from factor f to node $n \in \eta_f$ and by $\eta_n = \{f \in \mathcal{F} | n \in f\}$ the factors involving node n. The iterative update equations are then:

$$\mu_{n \to f}(x_n) = \psi_n(x_n) \prod_{\xi \in \eta_n \setminus \{f\}} \mu_{\xi \to n}(x_n) \quad (2)$$

$$\mu_{f \to n}(x_n) = \int \psi_f(x_{\eta_f}) \prod_{\xi \in \eta_f \setminus \{n\}} \mu_{\xi \to f}(x_\xi) \mathrm{d}x_{\eta_f \setminus \{n\}} \quad (3)$$

where the integral in Eq. (3) is taken over the appropriate measure, *i.e.*, a counting measure for discrete components, and a Lebesgue measure for continuous components. At convergence, the marginal distribution at a node n can be calculated as:

$$p(x_n) \propto \psi_n(x_n) \prod_{\xi \in \eta_n} \mu_{\xi \to n}(x_n) \quad (4)$$

In the case of a Markov chain, if all nodes are jointly Gaussian, then BP is equivalent to a Kalman smoother.

Similarly, if all nodes are discrete, then BP is equivalent to inference on an HMM using the forwards-backwards algorithm. BP extends each of these algorithms from chains to trees.

BP may be applied to loopy graphs (so-called *loopy belief propagation*). Practically, this simply involves repeated application of Eqs. (2) and (3) until convergence occurs (*i.e.*, until the maximum error between subsequent messages is less than a pre-set threshold). Unfortunately, this is neither guaranteed to converge to the right answer, nor to converge at all, although remarkable performance has been demonstrated in various applications [14]. Conceptually, one can always convert an arbitrary loopy graph to a tree by merging nodes (so-called *junction tree* representations [8]), but in practical problems, the dimensionality of the agglomerated variables may be prohibitive.

BP may also be applied to different probability distributions. In a graph involving a combination of discrete and Gaussian nodes and factors, all continuous messages will assume the form of Gaussian mixtures. Particle BP [15] extends particle filtering methods from Markov chains to general graphical models.

3 Graphical model data association

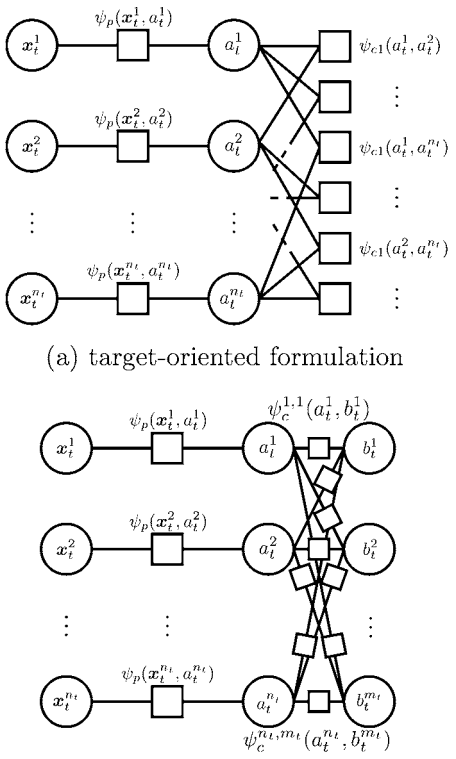
In this section, we construct graphical model formulations of the data association problem described in Section 2.1, thus permitting us to solve data association problems using the machinery described in Section 2.2. The formulations are shown in Fig. 1. In the following sections, we define the factors and write equivalent forms of Eq. (1) which respect the corresponding graph structures.

We do not include graph nodes for measurements in any of the graphs in Fig. 1. Since they are known at run-time, they do not need to be explicitly represented as nodes. Rather, the graph directly represents the conditional dependency structure, conditioned on the measurements, and the measurements appear as parameters in the factors in the graph.

3.1 Target oriented

First, we study the target oriented formulation, shown in Fig. 1(a). The constraint that each target gives rise to at most one measurement is implicit in the alphabet of the association variables a_t^i. The constraint that each measurement corresponds to at most one target is explicitly enforced by the consistency constraints $\psi_{c1}(a_t^i, a_t^j)$. Assuming that the prior distribution of $\boldsymbol{X}_t$ factorises into a product of terms in each target $\boldsymbol{x}_t^i$, we can write the Bayes update using Eq. (1) as

$$p(\boldsymbol{X}_t, \boldsymbol{a}_t | \boldsymbol{Z}^t) \propto \prod_{i=1}^{n_t} \left(\psi_p(\boldsymbol{x}_t^i, a_t^i) \prod_{j=i+1}^{n_t} \psi_{c1}(a_t^i, a_t^j) \right) \quad (5)$$



(a) target-oriented formulation

(b) hybrid formulation

Figure 1: Target-oriented and hybrid formulations of the data association problem using graphical models.

where

$$\psi_p(\boldsymbol{x}_t^i, a_t^i) = \begin{cases} [1 - P_d(\boldsymbol{x}_t^i)] p(\boldsymbol{x}_t^i | \boldsymbol{Z}^{t-1}), & a_t^i = 0 \\ \frac{P_d(\boldsymbol{x}_t^i) p(\boldsymbol{z}_t^{a_t^i} | \boldsymbol{x}_t^i) p(\boldsymbol{x}_t^i | \boldsymbol{Z}^{t-1})}{\lambda_{\text{fa}}(\boldsymbol{z}_t^{a_t^i})}, & a_t^i \neq 0 \end{cases} \quad (6)$$

$$\psi_{c1}(a_t^i, a_t^j) = \begin{cases} 0, & a_t^i = a_t^j \neq 0 \\ 1, & \text{otherwise} \end{cases} \quad (7)$$

This expression respects the graph structure of Fig. 1(a), *i.e.*, the expression is made up of factors that appear in the graph, involving variables to which the respective factors in the graph are connected.

If the consistency constraint factors $\psi_{c1}(a_t^i, a_t^j)$ are omitted from Fig. 1(a), the resulting association weights (*i.e.*, the marginal probabilities of a_t^i) are equivalent to the weights calculated by Probabilistic Data Association (PDA) [16]. Armed with our graphical formulation, the JPDA weights can be calculated by applying classical graph inference methods. As gating will generally restrict each target to associate with a small number of measurements, consistency constraint factors $\psi_{c1}(a_t^i, a_t^j)$ only need to be incorporated between targets (i, j) that may possibly associate with the same measurement. The resulting sparsity can be exploited by the junction tree method [8] to calculate the exact marginal probabilities of the association weights a_t^i in the presence of the consistency constraints. This can be implemented very easily using freely available

libraries such as libDAI [17], taking as an input the PDA weights and the constraint tables $\psi_{c1}(a_t^i, a_t^j)$; the result is quite similar to the fast mutual exclusion algorithm of [18, 19]. The disadvantage of the method is that its complexity is exponential in the width of the junction tree. Consequently, its computational complexity is problematic when many targets may share observations.

Alternatively, and somewhat naïvely, we may simply apply loopy BP. As we will see (in Section 4.1), the performance obtained by applying this approximate method to the present graph is somewhat disappointing. This is not surprising, given that *all* association variables are connected to each other, such that they form a large clique. The hybrid model described in Section 3.2 obtains much better performance with loopy BP.

3.2 Hybrid

The hybrid formulation is shown in Fig. 1(b). In this case, the constraints that each measurement corresponds to at most one target and each target gives rise to at most one measurement are both implicit in the alphabets of the respective association variables, a_t^i and b_t^j. The constraint factors $\psi_c^{i,j}(a_t^i, b_t^j)$ enforce consistency of the representations, *i.e.*, that the redundant association variables a_t^i and b_t^j describe the same association configuration. The definition of the factor $\psi_p(\boldsymbol{x}_t^i, a_t^i)$ is unchanged from Eq. (6), while the new constraint factor $\psi_c^{i,j}(a_t^i, b_t^j)$ is

$$\psi_c^{i,j}(a_t^i, b_t^j) = \begin{cases} 0, & a_t^i = j,\ b_t^j \neq i \text{ or } b_t^j = i,\ a_t^i \neq j \\ 1, & \text{otherwise} \end{cases} \tag{8}$$

so that we can write Eq. (1) as

$$p(\boldsymbol{X}_t, \boldsymbol{a}_t | \boldsymbol{Z}^t) \propto \prod_{i=1}^{n_t} \left(\psi_p(\boldsymbol{x}_t^i, a_t^i) \prod_{j=1}^{m_t} \psi_c^{i,j}(a_t^i, b_t^j) \right) \tag{9}$$

When an optimal inference algorithm is applied to this model, the result will be identical to the exact result obtained from the target oriented model; the formulation enforces constraints in a different (but equivalent) manner. However, as we will see, approximate algorithms may obtain quite different results.

The recent work [20] proves that loopy max-product BP applied to a minor variation of the hybrid graph in Fig. 1(b) is guaranteed to converge to the MAP assignment.[2] While this guarantee does not extend to the use of loopy sum-product BP in finding marginal association weights, the experiments detailed in Section 4.1 demonstrate its remarkable performance when applied to the hybrid formulation.

[2] assuming uniqueness of the optimum.

3.3 Multiple scans/sensors

The formulations described so far have concentrated on the problem that exists within a single time step. Extension to multiple time steps and/or multiple sensors simply involves a replication of the nodes for each time step, connecting the target states over time with factors that encode the Markov transition kernel.[3] In many tracking problems, the resulting graph will be mixed discrete-Gaussian, so that the messages sent between continuous nodes (and from discrete to continuous) are in the form of Gaussian mixtures. In our experiments, we approach this by marginalising over the continuous target states. When marginalising variables in a graphical model, we must connect all of the variables' neighbouring factors to all of those factors' other neighbours. The result is a purely discrete graph which contains for each target an augmented association variable $\boldsymbol{a}^i = (a_{t-s+1}^i, \ldots, a_t^i)$ that hypothesises a sequence of associations over all s scans in consideration. These augmented association variables are connected to sets of measurement association variables for each scan (using the hybrid formulation of Section 3.2). In a sense this is equivalent to using Gaussian mixture messages (as there is a bijection between augmented target association variable values and Gaussian mixture components), but options for message simplification are somewhat more limited (essentially, only pruning may be used). Experiments with formulations that either directly use Gaussian mixture messages, or so-called *weak marginalisation* [10] are a subject of future work.

Max-product BP may also be performed on this graph, resulting in an approximate algorithm for multi-dimensional assignment. This is similar in concept to [21], although the formulation therein assumes an additive decomposition of association weights (as a sum of pairwise distances), rather than the standard association model that we utilise.

4 Experiments

We consider a scenario in which a group of n_t targets travel in grid formation over a 2D planar field with equidistant spacing between adjacent targets. The state dynamics and target-derived measurement equations are

$$\boldsymbol{x}_t^i = \mathbf{F}\boldsymbol{x}_{t-1}^i + \boldsymbol{w}_t^i, \quad \boldsymbol{z}_t^i = \mathbf{H}\boldsymbol{x}_t^i + \boldsymbol{v}_t^i$$

where the state of each target $\boldsymbol{x}_t^i$, consists of position and velocity in two dimensions, $\boldsymbol{w}_t^i \sim \mathcal{N}\{\mathbf{0}, \mathbf{Q}\}$ and $\boldsymbol{v}_t^i \sim \mathcal{N}\{\mathbf{0}, \mathbf{R}\}$ are i.i.d. dynamics and measurement noise processes respectively, $\mathbf{H} = [1\ 0] \otimes \mathbf{I}_2$, $\mathbf{R} = \mathbf{I}_2$, $\mathbf{F} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \otimes \mathbf{I}_2$ and $\mathbf{Q} = q \begin{bmatrix} T & T^2/2 \\ T^2/2 & T^3/3 \end{bmatrix} \otimes \mathbf{I}_2$.

[3] For the multiple sensor case, sets of association nodes for each sensor are connected to the same set of target state nodes. The target state prior $p(\boldsymbol{x}_t^i | \boldsymbol{Z}^{t-1})$ that is incorporated into Eq. (6) is only included in the $\psi_p(\boldsymbol{x}_t^i, a_t^i)$ factor for the first time step/sensor.

The sample period $T = 1$ sec, $q = 10^{-2}$, and $\mathbf{I}_2$ is a 2×2 identity matrix. Targets are detected with probability $P_d = 0.6$ (unless otherwise stated), and false alarms are distributed uniformly in a 100×100 region (which covers all targets), arriving according to a Poisson process where the expected total number is equal to 0.1 (unless otherwise stated).[4] State estimates and covariances are initialised with values consistent with those that would be obtained through a random process involving five measurements at consecutive time steps with no association uncertainty. For each experiment, 1000 Monte Carlo runs were performed. The libDAI [17] implementation of loopy BP was used with parallel updates, a convergence threshold of 10^{-6} and a maximum of 1000 iterations.

4.1 Single scan

4.1.1 Marginal association probability errors

We first consider a scenario involving six targets[4] in a regular 2×3 grid. We vary the spacing between the targets and examine the behaviour of our approximate algorithms, comparing the marginal association weights calculated to the exact JPDA weights. We examine the average maximum association weight error per target for the target-oriented formulation (Section 3.1) and the hybrid formulation (Section 3.2). We also show the error in the PDA weights (*i.e.*, the weights calculated ignoring consistency constraints) for reference, as a gauge of the degree of interaction between targets, and hence of the difficulty of the scenario.

Results of experiments with probability of detection $P_d = 0.3$, 0.6 and 0.9 are shown in Fig. 2(a-c,f-h,k-m). The plots of the average maximum association probability error per target in (a-c) demonstrate that, while the target-oriented formulation of Section 3.1 (dashed line) commits errors of the same order of magnitude as PDA (dot-dashed line), the errors in the weights calculated by the hybrid formulation of Section 3.2 (solid line) are quite small. For $P_d = 0.9$ the errors are reasonably small (≤ 0.083), but for lower P_d, they are remarkably small (≤ 0.026 for $P_d = 0.6$, and ≤ 0.006 for $P_d = 0.3$). The results of experiments with an increased false alarm rate (from $\lambda_{\text{fa}} = 0.1$ to $\lambda_{\text{fa}} = 1$ and $\lambda_{\text{fa}} = 2$) with $P_d = 0.9$ are shown in Fig. 2(d-e,i-j,n-o). The results show that the accuracy of the approximate weights calculated by the hybrid formulation improves as the false alarm rate increases (error ≤ 0.072 for $\lambda_{\text{fa}} = 1$ and ≤ 0.068 for $\lambda_{\text{fa}} = 2$).

One factor which acts to reduce coupling in the graph (and hence reduce errors) is the "missed detection" events, under which there is no competition between targets for measurements. These events are more probable when P_d is lower, and when λ_{fa} is higher. Accordingly, as demonstrated in the results, the BP-based algorithm can be expected to produce better approximations of association weights in challenging tracking environments, *i.e.*, those involving lower probability of detection and/or higher false alarm rates.

The plots in Fig. 2(f-j) show the percentage of cases in which loopy BP converges, while (k-o) show the average number of iterations (excluding non-convergent cases). The results show that, while the target-oriented formulation exhibits significant difficulties with non-convergence, the hybrid formulation converges very reliably (100% and $\geq 97.6\%$ of trials) for $P_d = 0.3$ and 0.6. While convergence is less reliable ($\geq 46.4\%$ of trials) for $P_d = 0.9$, the average error differs little between cases that converge and those that do not, so the oscillation that is occurring would appear to be between reasonably good solutions. With higher false alarm rates (and $P_d = 0.9$), the hybrid formulation converges very reliably ($\geq 99.6\%$ and 100% of trials for $\lambda_{\text{fa}} = 1$ and $\lambda_{\text{fa}} = 2$ respectively). No attempt was made to employ BP convergence aids such as damped updates.

Given the significant performance advantage of the hybrid formulation over the target-oriented formulation, the subsequent experiments employ it exclusively. The marked difference in the performance of the target oriented formulation and the hybrid formulation raises the question of whether there may be structures involving further redundancy that provide additional increases in performance; this is a topic of future study.

4.1.2 Scalability of loopy BP method

Our second experiment examines the scalability of our proposed method by calculating association weights for a square grid of between 2×2 and 10×10 targets (*i.e.*, between four and 100 targets), with a spacing of four units. Gating is applied conservatively such that a factor is not incorporated between a target and a measurement if the PDA weight is less than 10^{-6} (the alphabet of the measurement association variable is constrained accordingly). In the 10×10 target case, the maximum number of measurements to which a target is connected is 11.3 on average (averaging over the 1000 Monte Carlo trials), and the targets are connected to 6.3 measurements on average (averaging over targets and Monte Carlo trials). This level of connectivity across a dense grid is unable to be addressed by any exact method.

Fig. 2(p) shows the average True Measurement Probability (TMP), *i.e.*, the marginal weight of the correct association for the target, for different problem sizes (solid line), and the average number of iterations to convergence (solid line with crosses). The diagram shows that the quality of the solution suffers little degradation as the problem size increases, and that the number of iterations required grows at a relatively slow rate. Asymptotically, we expect that this will level out to a constant value as the number of targets exceeds the graph connectivity, *e.g.*, since associations of distant

[4] This relatively small number was chosen to permit calculation of the exact weights (for comparison) in reasonable time.

targets are of little consequence. Loopy BP converged in all trials and all network sizes.

The average run time for the 100 target scenario was 19.4 sec. Without gating, there are $n_t \times m_t$ factors, each of which involves computation (when implemented using a general-purpose inference library) of $(n_t + 1) \times (m_t + 1)$ elements. Our naïve implementation with conservative gating involves (on average) 793 factors, each of which involves (on average) over 4900 elements. The number of factors can be reduced by using a less conservative gating strategy (10^{-3} is a more common threshold). The structure of the constraint matrices may also be exploited to obtain faster run times; specifically, each target-measurement and measurement-target message involves only two different values. This is not exploited in our implementation, which utilises the general-purpose libDAI library, such that all values are calculated independently. We expect that an optimised, purpose specific implementation could speed computation by two orders of magnitude on the same hardware and using the same conservative gating strategy. In an optimised implementation (neglecting gating), each iteration of target association variable to measurement association variable message calculations will require (in total) $2n_t m_t^2$ floating point operations, while each iteration of measurement association variable to target association variable message calculations will require $2n_t^2 m_t$ operations. In addition to these optimisations, the message passing structure of BP lends itself to parallelisation, allowing for easy exploitation of multi-core architectures (*e.g.*, [22]).

4.2 Multiple scans

4.2.1 Marginal association weights

The experiment setup described in Section 4 was also applied to scenarios involving two, three and four scans of measurements (with six targets in a 2×3 grid, spaced by four units). We apply loopy BP to the entire graph (spanning multiple time steps) as described in Section 3.3, comparing results to JPDA,[5] in which marginal weights for each target are calculated exactly for each time step, and then propagated forward approximating the past joint association weights as the product of the marginal weights for each target. The computational complexity of an optimised implementation of this is approximately $n_t n_h \sum_{k=t-s+1}^{t} m_t$ and $n_t^2 \sum_{k=t-s+1}^{t} m_t$ floating point operations (in total) for each alternating iteration, where n_h is the number of (single target) association hypotheses per target. We cannot solve for the optimal augmented (over time) weights in a problem of this size. BP failed to converge in only 0, 1, 1 and 3 cases out of 1000 in the 1, 2, 3 and 4 scan experiments respectively. The average number of iterations required was 24, 27, 31 and 28 respectively (excluding non-convergent cases).

Our results are shown in Fig. 2(q-t). The x and y axes plot TMP in the final time step for loopy BP and JPDA respectively. Intuitively, we expect that higher TMP values represent better performance. Points to the right of/below the $x = y$ line on the plots represent cases in which loopy BP yields a better result, while points above/to the left of the $x = y$ line represent cases in which JPDA yields a better result. The plots for the single scan case (Fig. 2(q)) reveal more about the nature of the errors committed by loopy BP in the experiments in the previous section. In this case, the JPDA weights represent the exact marginal weights for the scenario. In comparison, the BP result tends to apply higher weight to the true measurement when its weight is higher, and lower weight to the true measurement when its weight is lower. The overall average error is consistent with the result shown in Fig. 2(b). In the 2, 3 and 4 scan experiments in Fig. 2(r-t), there are progressively fewer cases in which BP assigns a higher weight to the true measurement, and a small number of cases appear in which BP assigns very low weight (*e.g.*, ≈ 0.01) to the true hypothesis when JPDA assigns a greater value (*e.g.*, $\approx 0.1 - 0.2$). Consequently, although BP yields a slightly higher ($< 1\%$) mean TMP in all experiments, it generally exhibits a slightly lower ($< 0.8\%$) mean log TMP, which is arguably the correct measure for comparison. The erroneously low weights are of practical importance as they would likely lead to pruning of the correct hypothesis in a small proportion of cases.

It is plausible that BP might be able to provide better performance than JPDA over multiple time steps as it is able to implicitly represent correlation between target association events in earlier time steps. This improvement is not evident in this experiment. Rather it seems that the strong loops that arise due to coupling of time steps causes occasional anomalous behaviour. Future work is required on this topic to examine the causes of these effects, identify cases in which problems are likely to occur, and develop alternative methods for application in these cases.

4.2.2 Multi-dimensional assignment

As mentioned in Section 3.3, the max-product variant of loopy BP can also be applied to find an approximate MAP solution, *i.e.*, an approximate solution to multi-dimensional assignment problems. To demonstrate the potential of the algorithm, we applied max-product BP to the problem described in the previous section, and compared to the optimal solution obtained using the Matlab Optimization Toolbox. The results for $10,000$ Monte Carlo trials are summarised in Table 1.

[5] Whereas standard JPDA approximates the posterior as a unimodal Gaussian, we retain the full Gaussian mixture for each target with the corresponding weights. Mixture reduction is obviously necessary in practice; in this work we retain the full representation in order to concentrate on the problem at hand, *i.e.*, approximate inference of association weights.

# Scans	Conv	Feas	Optimal	# Iter
1	9995	10000	10000	9
2	9926	9938	9929	10
3	9941	9953	9935	10
4	9942	9953	9937	11

Table 1: Results of loopy max-product BP applied to multi-dimensional assignment problems. Columns show number of scans (*i.e.*, number of dimensions in assignment problem minus 1), number of Monte Carlo trials (out of $10,000$) in which BP converged (we expect that all simulations in the one-scan case would have converged if iteration was permitted to continue), number of trials in which BP produced a feasible answer (*i.e.*, one which obeys the consistency constraints), number of trials in which BP produced an optimal solution (*i.e.*, an objective within 10^{-6} of the integer programming solution), and average number of iterations required for BP to converge (excluding non-convergent cases).

These results demonstrate the potential of loopy BP as a viable alternative solution for multi-dimensional assignment problems. For example, even in a four scan (5D) assignment problem, the optimal solution was found in 99.4% of cases. Our future work on this problem includes testing the formulation in more stressing scenarios, and applying Tree Re-weighted Max Product (TRMP) [23], which can improve performance over max product BP, and provide a "certificate of optimality" in some circumstances.

5 Discussion

The experiments presented in Section 4 demonstrate the remarkable performance of the hybrid graphical model formulation of Section 3.2 for data association problems. The experiments were designed to provide challenging conditions involving a range of interaction strengths and a large number of targets. The performance of the graphical model-based method is clearly impressive, especially in challenging environments, *i.e.*, problems with low P_d and/or high false alarm rates. This performance, coupled with the low order polynomial complexity of the method, demonstrate that loopy BP is a very attractive method for both calculation of marginal association weights and multi-dimensional assignment.

Our future work includes comparing to existing approximate methods, incorporating varying and uncertain target counts, and employing and/or developing advanced inference methods to address the small proportion of cases where performance is problematic.

6 Acknowledgements

The authors thank Prof Alan Willsky and Prof Müjdat Çetin for many helpful discussions during development of the formulations in Section 3. Our experiments utilised the graphical model inference engine libDAI [17].

References

[1] L. Chen, M. J. Wainwright, M. Çetin, and A. S. Willsky, "Multitarget-multisensor data association using the tree-reweighted max-product algorithm," in *Proc SPIE Signal Processing, Sensor Fusion, and Target Recognition*, vol. 5096, August 2003, pp. 127–138.

[2] L. Chen, M. Çetin, and A. S. Willsky, "Distributed data association for multi-target tracking in sensor networks," in *Proc. Eighth International Conference on Information Fusion*, July 2005.

[3] L. Chen, M. J. Wainwright, M. Çetin, and A. S. Willsky, "Data association based on optimization in graphical models with application to sensor networks," *Mathematical and Computer Modelling*, vol. 43, no. 9–10, pp. 1114–1135, 2006.

[4] K.-C. Chang and Y. Bar-Shalom, "Joint probabilistic data association for multitarget tracking with possibly unresolved measurements and maneuvers," *IEEE Transactions on Automatic Control*, vol. 29, no. 7, pp. 585–594, July 1984.

[5] A. B. Poore and S. Gadaleta, "Some assignment problems arising from multiple target tracking," *Mathematical and Computer Modelling*, vol. 43, no. 9–10, pp. 1074–1091, 2006.

[6] P. Horridge and S. Maskell, "Searching for, initiating and tracking multiple targets using existence probabilities," in *Proc. Twelfth International Conference on Information Fusion*, July 2009, pp. 611–617.

[7] R. L. Streit and T. E. Luginbuhl, "Probabilistic multihypothesis tracking," Naval Undersea Warfare Center Division, Newport, RI, Tech. Rep., 1995.

[8] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. San Francisco, CA: Morgan Kaufmann, 1988.

[9] M. I. Jordan, "Graphical models," *Statistical Science*, vol. 19, no. 1, pp. 140–155, 2004.

[10] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter, *Probabilistic Networks and Expert Systems: Exact Computational Methods for Bayesian Networks*. New York, NY: Springer, 2007.

[11] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[12] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb 1989.

[13] F. Kschischang, B. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, Feb 2001.

[14] K. P. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: An empirical study," in *Uncertainty in Artificial Intelligence*, 1999.

[15] A. Ihler and D. McAllester, "Particle belief propagation," *Journal of Machine Learning Research*, vol. 5, pp. 256–263, 2009.

[16] Y. Bar-Shalom and E. Tse, "Tracking in a cluttered environment with probabilistic data association," *Automatica*, vol. 11, no. 5, pp. 451 – 460, 1975.

[17] J. M. Mooij, et al, "libDAI 0.2.4: A free/open source C++ library for Discrete Approximate Inference," http://www.libdai.org/, 2010.

[18] P. Horridge and S. Maskell, "Real-time tracking of hundreds of targets with efficient exact JPDAF implementation," in *Proc. Ninth International Conference on Information Fusion*, July 2006, pp. 1–8.

[19] S. Maskell, M. Briers, and R. Wright, "Fast mutual exclusion," in *Proc SPIE Signal and Data Processing of Small Targets*, vol. 5428, 2004, pp. 526–536.

[20] M. Bayati, D. Shah, and M. Sharma, "Max-product for maximum weight matching: Convergence, correctness, and LP duality," *IEEE Transactions on Information Theory*, vol. 54, no. 3, pp. 1241–1251, March 2008.

[21] H. Zhu, C. Han, and C. Li, "Graphical models-based track association algorithm," in *Proc. Tenth International Conference on Information Fusion*, July 2007, pp. 1–8.

[22] C. Webers, K. Gawande, A. Smola, S. V. N. Vishwanathan, S. Günter, C. H. Teo, J. Q. Shi, J. McAuley, L. Song, and Q. Le, "Elefant user manual (revision 0.1)," NICTA, Tech. Rep., 2009, http://elefant.developer.nicta.com.au.

[23] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Foundations and Trends in Machine Learning*, vol. 1, no. 1–2, pp. 1–305, 2008.

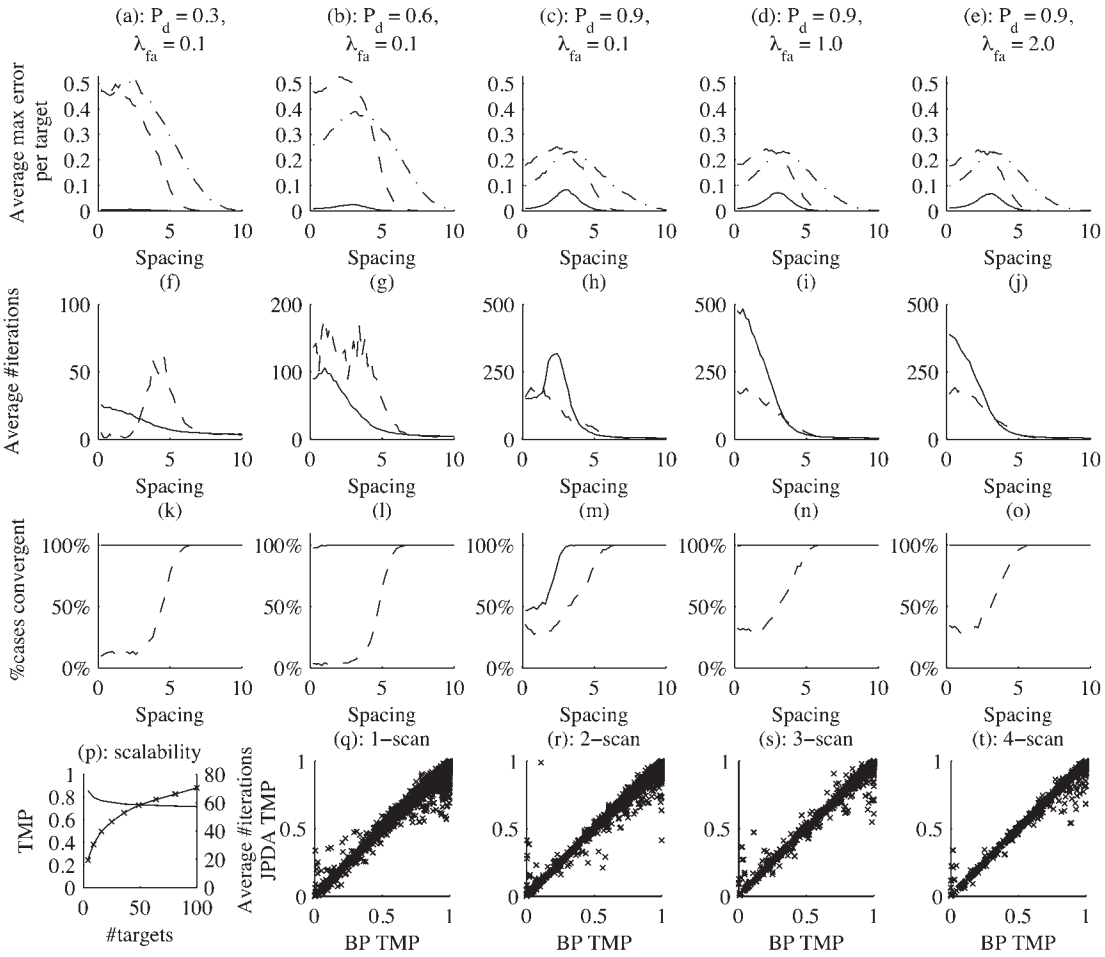


Figure 2: Results of computational experiments. (a-o) show the average maximum association weight error per target, the average number of iterations and percentage of cases in which loopy BP converges for the experiments in Section 4.1.1. The solid line shows the loopy BP hybrid formulation, the dashed line shows the loopy BP target-oriented formulation, while the dot-dashed line shows the PDA weights (neglecting consistency constraints). (p) shows the average marginal probability of the true measurement (or True Measurement Probability, TMP) with an unmarked line and average number of iterations with a cross marked line for the scalability experiments (Section 4.1.2). (q-t) show TMP for loopy BP (x axis) and JPDA (y axis) for experiments involving 1, 2, 3 and 4 scans of measurements (Section 4.2.1).